Performance Analysis of Deep Learning based Human Activity Recognition Methods

Mst. Farzana Aktter[a], Md. Anwar Hossain[a,*], Sohag Sarker[a], AFM Zainul Abadin[a] and Mirza AFM Rashidul Hasan[b]
[a]*Pabna University of Science and Technology, Bangladesh*
[b]*University of Rajshahi, Bangladesh*

Abstract

Human Activity Recognition (HAR) is one of the most important branches of human-centered research activities. Along with the development of artificial intelligence, deep learning techniques have gained remarkable success in computer vision. In recent years, there is a growing interest in Human Activity Recognition systems applied in healthcare, security surveillance, and human motion-based activities. A HAR system is essentially made of a wearable device equipped with a set of sensors (like accelerometers, gyroscopes, magnetometers, heart-rate sensors, etc.). Different methods are being applied for improving the accuracy and performance of the HAR system. In this paper, we implement Artificial Neural Network (ANN), and Convolutional Neural Network (CNN) in combination with Long Short-term Memory (LSTM) methods with different layers and compare their outputs towards the accuracy in the HAR system. We compare the accuracy of different HAR methods and observed that the performance of our proposed model of CNN 2 layers with LSTM 1 layer is the best.

Keywords: *Human Activity Recognition (HAR), Artificial Neural Network (ANN), Convolutional Neural Network (CNN), Long Short-term Memory (LSTM)*

1. Introduction

Human activity recognition (HAR) is a classification task for recognizing human movements. It is a technique that trains a supervised learning model to recognize activity performed by the human body. Human activity recognition can be performed using images, video, and sensor data. Recently, deep neural networks have been deployed for HAR in the context of activities of daily living using multichannel time-series. These time-series are acquired from body-worn devices, which are composed of different types of sensors [1]. The deep architectures process these measurements for finding basic and complex features in human corporal movements, and for classifying them into a set of human actions [2].

With the rapid development of information technology and the popularity of smart devices, it is easier to gather data that can describe human daily activities collected from various sensors, which are integrated into smart devices [3]. Apparently, instead of attaching a variety of cumbersome sensors to the user's body, people are more willing to accept portable, wearable, and multi-functional devices such

* Corresponding author. Tel.: +88 01717 330923
E-mail address: manwar.ice@pust.ac.bd

Manuscript History:
Received 1 April, 2022, Revised 28 April, 2022, Accepted 1 May, 2022, Published 31 October, 2022

https://doi.org/10.33736/jaspe.4639.2022

e-ISSN: 2289-7771

as smartphones and smart-watches, which also embed various sensors, for instance, accelerometers and gyroscopes. Therefore, more and more methods and software applications based on smartphone sensors are proposed and developed for human activity detection [4].

To reduce the impact of the diversity of human activity patterns in human activity recognition, we implement a Convolutional Neural Network (CNN)-based method that uses the raw accelerometer data collected from the smartphone accelerometer sensor, which has less computation, higher recognition accuracy, and higher flexibility and robustness. The six kinds of human activities we chose to be recognized were walking, jogging, sitting, standing, upstairs, and downstairs [4, 5].

For performance comparison, we implement Long Short-Term Memory (LSTM) recurrent networks with the model for System. LSTM decides based on current and previous data. Adding only a hidden layer could significantly improve the performance of the model we add a convolutional layer to extract features [5]. We then compare the performance of different models to determine their accuracy and smoothness.

2. Materials and methods

2.1. Methodology

In this section, we discuss the methodology of our work. As we implemented 5 methods using Artificial Neural Network, Long Short-term Memory, Convolutional Neural Network, and a modified combination of CNN-LSTM with different layers, we describe their structure and working procedure used in this paper.

2.1.1. Artificial Neural Networks (ANN) model

In our first method of the HAR system, an ANN with 1152(input layer) X 48 (hidden layer) X 24 (hidden layer) X 12 (hidden layer) X 6 (output layer) network structure is used. Here, dense layer is used in the thorough input-to-output layer to make every node fully connected and the Dropout layer to prevent overfitting. We use Rectified Linear Unit (ReLU) activation function in the input and hidden layers and the SoftMax activation function in the output layer. Adam Optimization technique is implemented in this network structure which is suited for problems with a large amount of data and parameters. Figure 1 shows the network architecture of the ANN model for the HAR system.

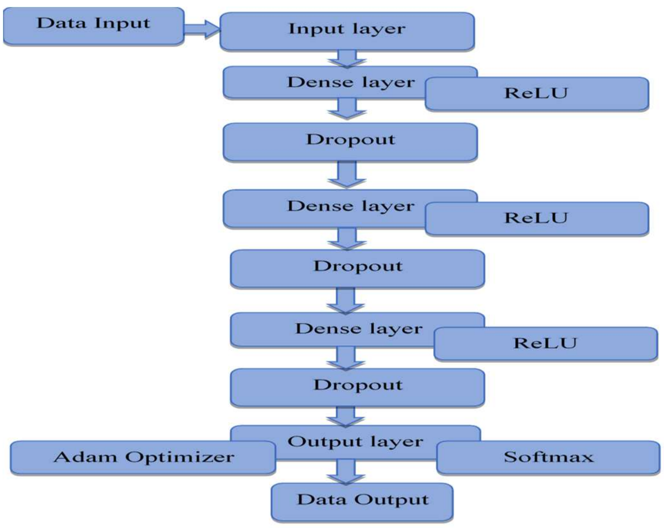


Figure 1: ANN architecture model

2.1.2. Long Short-term Memory (LSTM) model

In our next method, Long Short-term Memory (LSTM) is a recurrent neural network architecture. It can not only process single data points but also entire sequences of data which makes it suitable for large data processing like HAR. Also, in this method, we use dense layer and dropout layer, ReLU, and SoftMax activation function. Here too is the Adam optimization technique used. Figure 2 illustrates the basic structure of the LSTM network for the HAR system.

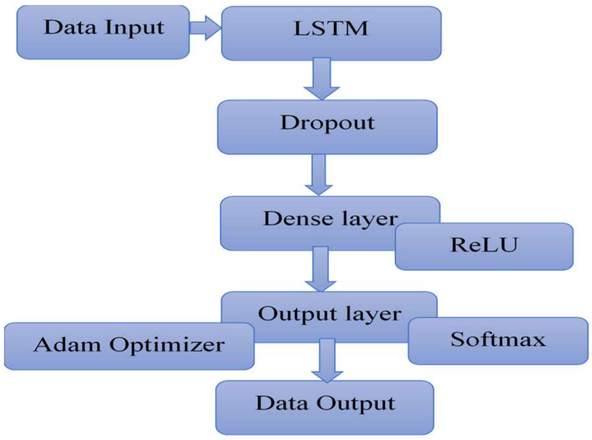


Figure 2: LSTM network architecture

2.1.3. Convolutional Neural Network with Long Short-term Memory model

CNN is a deep neural network that is used for feature extraction [6]. LSTM is a type of recurrent neural network, so it can remember past data. Its decisions are influenced by what it has learned from the past [7]. In this model, we combine CNN and LSTM. Both approaches have been reported to provide improved results in areas such as image processing, and voice recognition [7, 8]. In this model, we use two layers of CNN and one layer of LSTM. The CNN layers are followed by a max-pooling layer. It reduces the computational cost by reducing the number of parameters to learn and provides basic translation invariance to the internal representation [9]. The output of the max-pooling layer is then flattened to feed into the LSTM layer. Figure 3 shows the structural block diagram of this model.

2.1.4. Optimized CNN 2 layers with LSTM 1 layer

This is the optimized version of the previous model of CNN with LSTM architecture. Since model performance depends on hyperparameters, we can introduce some hyperparameters. In this model, we introduce batch normalization and dropout layer as hyperparameters. Batch normalization is used for training a deep neural network that standardizes the inputs to a layer for each mini-batch [10]. This also stabilizes the learning process. Dropout is a regularization technique that can reduce the chance of overfitting. We can set the rate of dropout we want to drop in a layer. Therefore, if we use batch normalization along with dropout, model performance will be better.

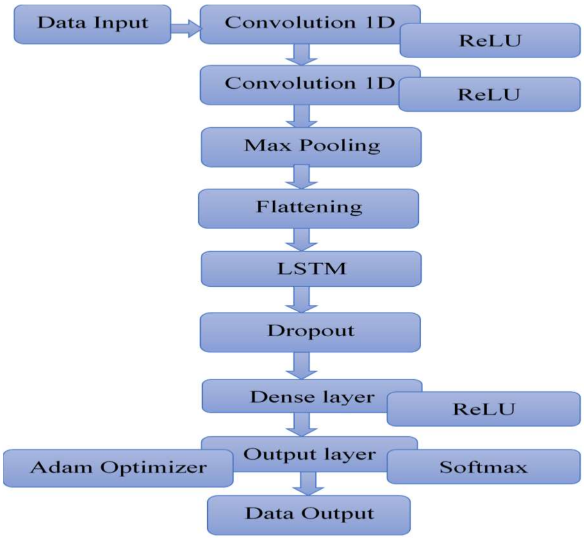


Figure 3: Two-layer CNN with single Layer LSTM

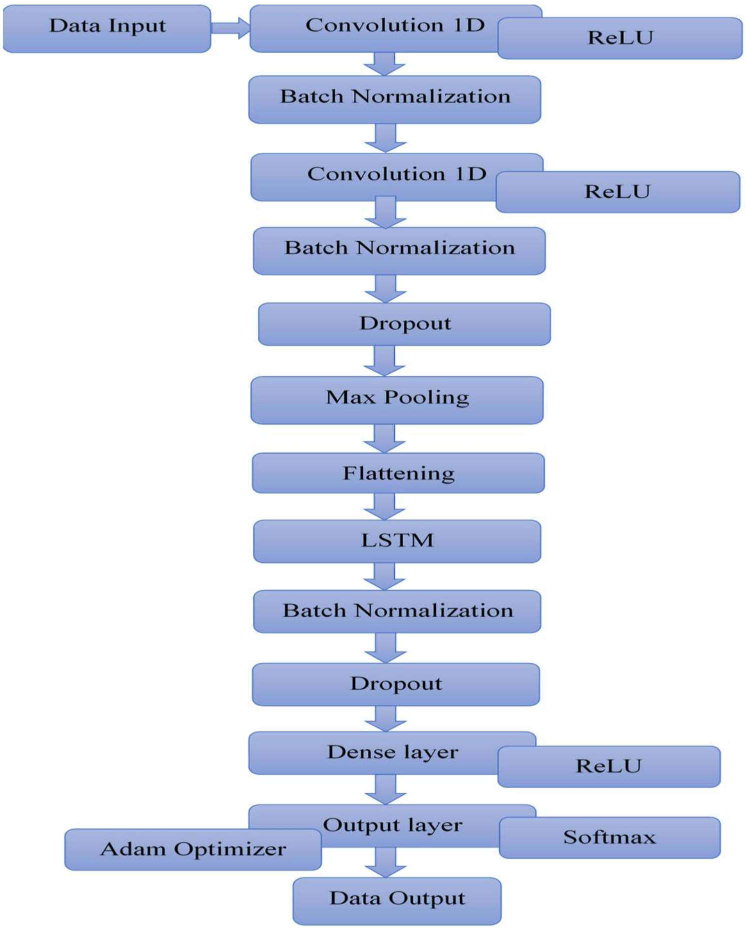


Figure 4: Optimized CNN 2 layers with LSTM 1 layer

2.1.5. CNN 4-layer with LSTM 2-layer

Figure 5 shows the block diagram of a combined network of CNN 4-layer with LSTM 2 layer. In this model, the convolutional and LSTM layer is increased to observe the performance variation from models 3 and 4.

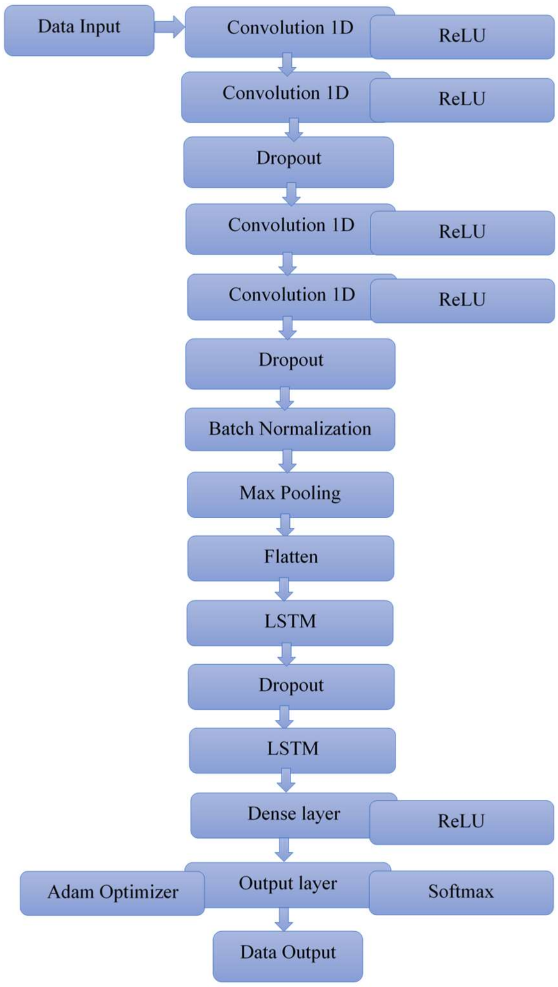


Figure 5: Optimized CNN 4-Layer with LSTM 2-Layer

2.2 Dataset

The methods described above are tested with a public data set UCI HAR Dataset collected from the website https://machinelearningmastery.com/cnn-models-for-human-activity-recognition-time-series-classification/. The experiments have been carried out with a group of 30 volunteers within an age bracket of 19-48 years. Each person performed six activities (WALKING, WALKING_UPSTAIRS, WALKING_DOWNSTAIRS, SITTING, STANDING, LAYING) wearing a smartphone (Samsung Galaxy S II) on the waist. Using its embedded accelerometer and gyroscope, 3-axial linear acceleration and 3-axial angular velocity were captured at a constant rate of 50Hz. The experiments have been video-recorded to label the data manually. The obtained dataset has been randomly partitioned into two sets, where 70% of the volunteers were selected for generating the training data and 30% for the test data

[11]. In statistical modeling, data is frequently split randomly 70-30 or 80-20 into train and test datasets, with training data being used to develop the model and its effectiveness being tested on test data.

3. Results and discussion

We evaluate different models on public datasets namely the UCI HAR dataset using two different ways: (1) cross-entropy and (2) confusion matrix. In this section, we compare the activity recognition performance of these models on this dataset. These experiments were evaluated in a google colab environment.

3.1. Evaluate using cross-entropy

Working with different neural network models, different results are obtained for accuracy and loss based on the model. The first model is ANN. For ANN architecture, the model train well, and the training loss is decreasing with the increasing number of epochs. As the number of epochs is increasing, this model can recognize human activities more accurately and validation accuracy increases. Figure-6 shows the comparative diagrams for the performance of the ANN model.

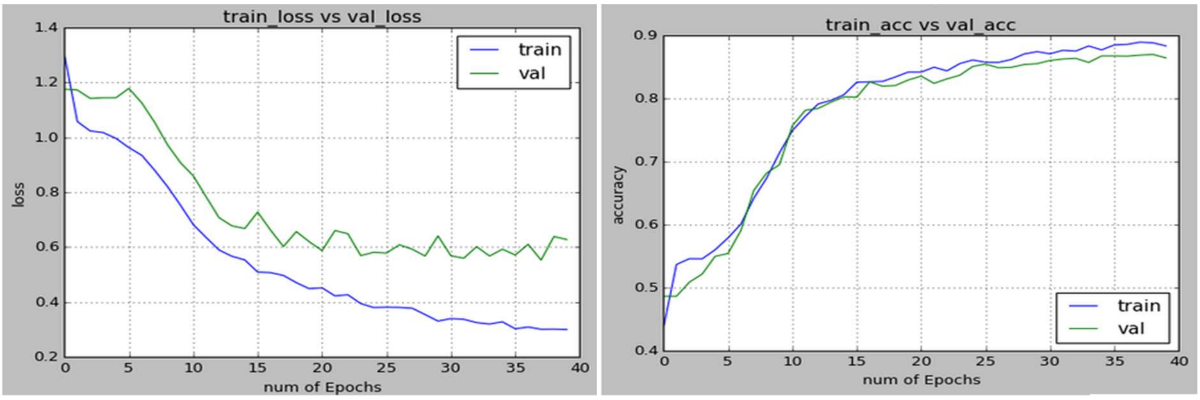


Figure 6: Diagram of train loss vs. validation loss and train accuracy vs. validation accuracy for ANN.

In our 2nd model of the LSTM 1 layer network, with increasing epochs, the training loss and validation loss decrease. As the model train well with more epochs, then validation accuracy increases. The comparative diagram of 1 layer LSTM is shown in Figure-7.

3rd method was CNN 2 layer with LSTM 1 layer. Here, we use two convolution layers with one LSTM layer. It provides the comparison graph as shown in Figure-8. With increasing the number of epochs, the model train well, so train loss is minimized. But validation loss is not minimized. Both the training accuracy and validation accuracy are increasing with the number of epochs increasing.

In the 4th method, batch normalization along dropout implement in CNN 2 layers with 1 LSTM layer that improves the performance of model 3. The performance of the model is greatly influenced by hyperparameters. [12,13]. The dropout layer is used to reduce the risk of overfitting [14] and batch normalization is used for a deep neural network with many layers to stabilize the learning process. From Figure 9, we can see the accuracy is high in comparison to model 3. Both the validation accuracy and train accuracy increase as the number of epochs increases.

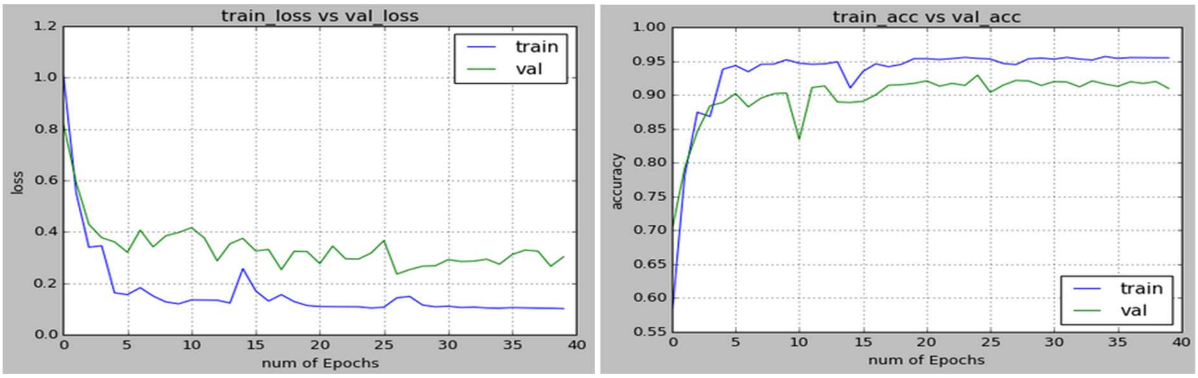


Figure 7: Diagram of train loss vs. validation loss and train accuracy vs. validation accuracy for LSTM 1 layer.

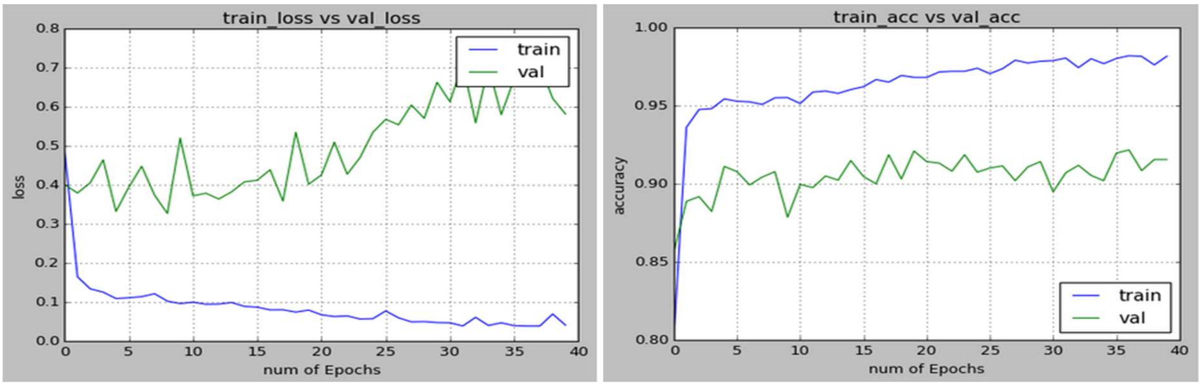


Figure 8: Diagram of train loss vs. validation loss and train accuracy vs. validation accuracy for CNN 1 layer with LSTM 2 layer.

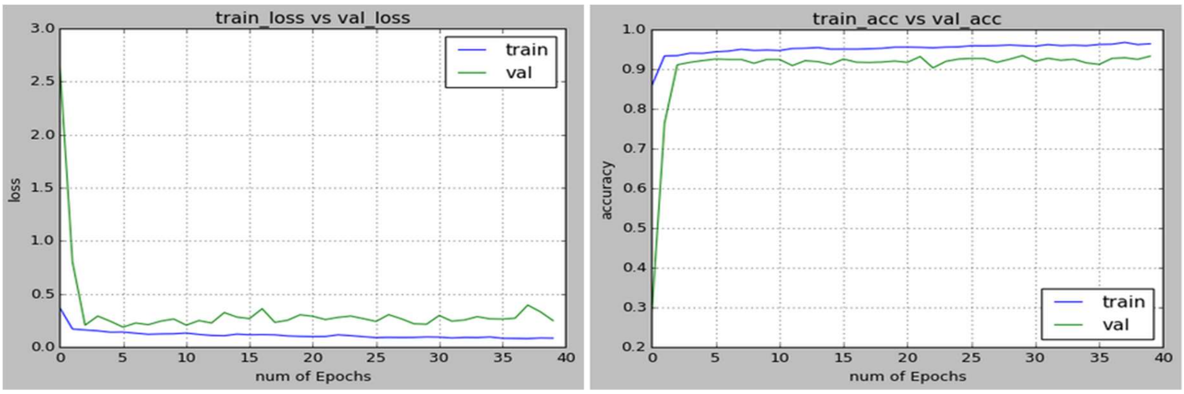


Figure 9: Diagram of train loss vs. validation loss and train accuracy vs. validation accuracy for Optimized CNN 1 layer with LSTM 2 layer.

The 5th model for the HAR system was 4 layers CNN with 2 layers LSTM. Figure -10 shows the performance graph for this model. This model also includes dropout and batch normalizations but the number of layers is increased. The performance of the model is degrading as compared to model 4.

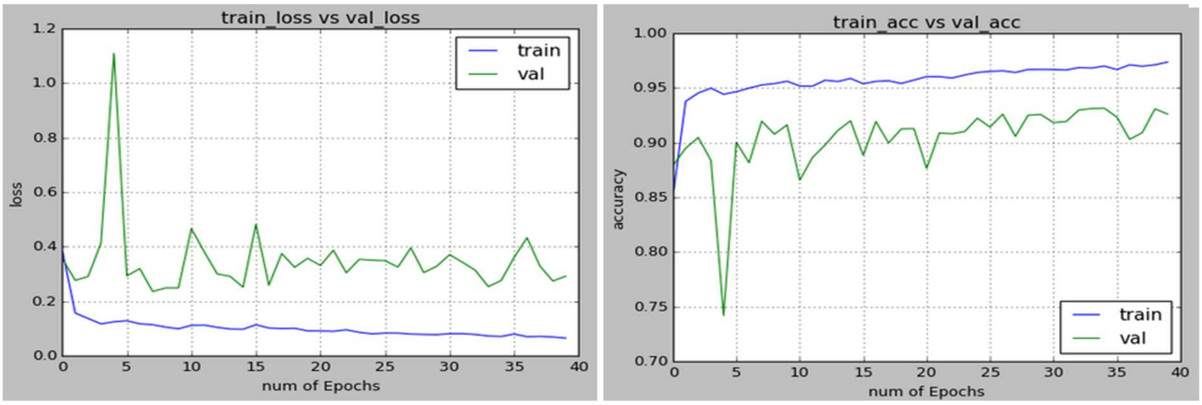


Figure 10: Diagram of train loss vs. validation loss and train accuracy vs. validation accuracy for CNN 4 layer with LSTM 2 layer.

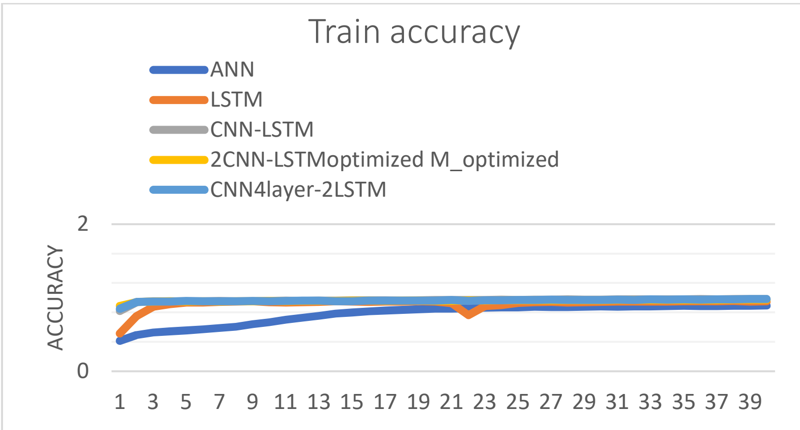


Figure 11: Comparative diagram of train accuracy.

In the Figure 11 comparative diagram of the train, accuracy is depicted. The test is conducted with 40 epochs from the data set. The variation is observed up to a certain epoch then the variation remains almost constant. For CNN4Layer-LSTM2 the variation is observed for more epochs than other modules.

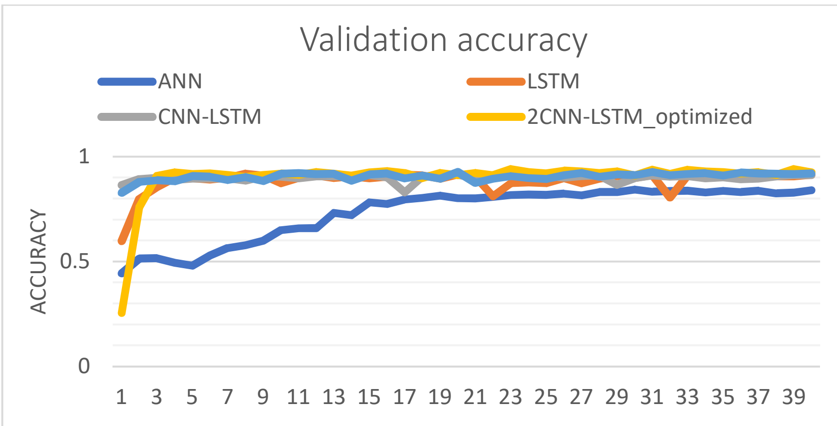


Figure 12: Comparative diagram of validation accuracy.

Figure 12 depicts a comparative diagram of Validation accuracy. In this diagram, we observe the variation with the increase of epoch. Several modules preserve the variation up to a certain epoch. Then the variations approach with an approximately constant value. For CNN4LAYER-LSTM2 the accuracy is observed with more variation. From both Figure-11 and Figure-12, we can see that performance of model 4 is better because we use batch normalization along dropout technique.

The second purpose of our paper is the comparison of the performance i.e. validation accuracy rate and trained accuracy rate of the implemented methods. The following table shows the validation accuracy of the data set of 40 epochs. The performance comparison of models on the UCI HAR dataset in terms of accuracy is summarised in Table-1. From this table, we can see, that the performance of our proposed model of optimized CNN 2 layers with LSTM 1 layer is the best.

Table 1: Accuracy rate of experimented models

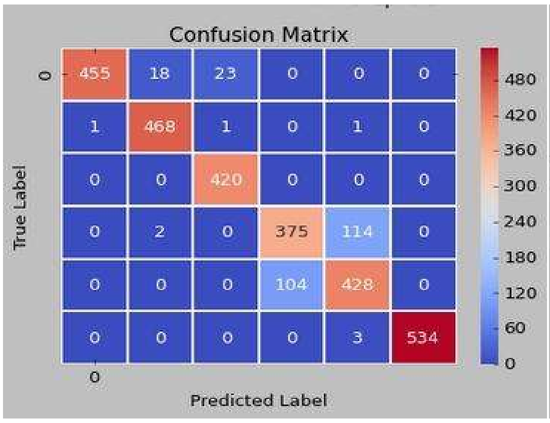
Model	Method of Test	Accuracy rate
Model 1	ANN	84.42%
Model 2	LSTM 1 layer	91.35%
Model 3	CNN 2 layer with LSTM 1 layer	91.25%
Model 4	Optimized CNN 2 layer with LSTM 1 layer	92.60%
Model 5	CNN 4 layer with LSTM 2 layer	91.89%

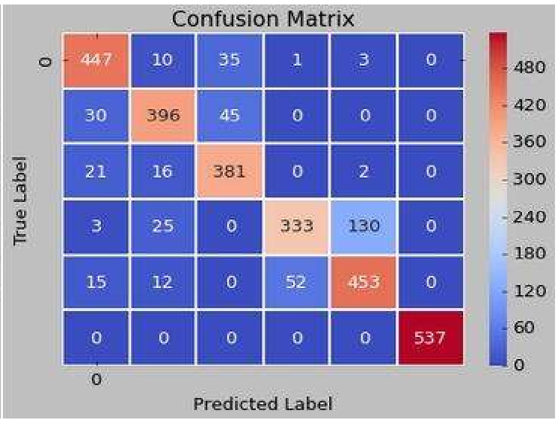
3.2. Evaluation using confusion matrix

We also observe our models using a confusion matrix. The confusion matrix shows predicted outputs with corresponding actual outputs. From the confusion matrix, we can see how accurately the model predicts output i.e. accuracy of the model. Figure 13 shows the confusion matrices of our models.

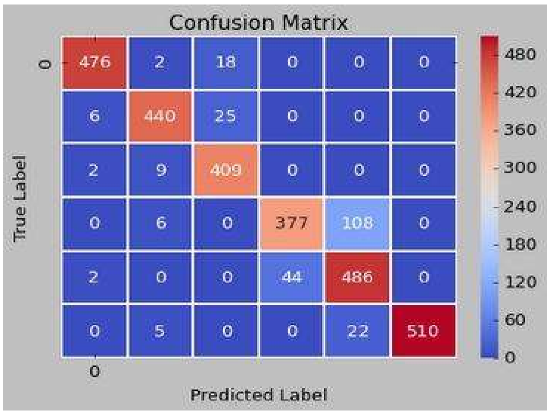
Table 2: Class-wise right(R) and wrong(W) cases for UCI HAR dataset over different models (Testing sample 2947)

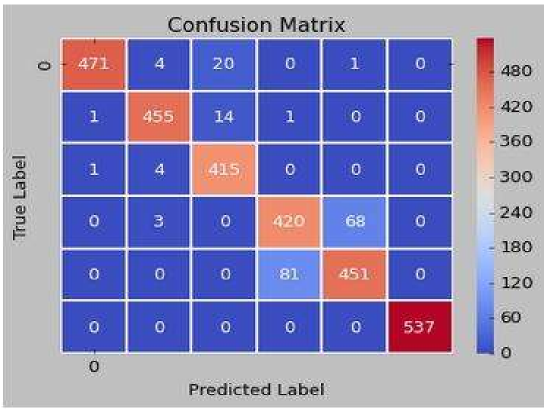
Class No	Class Name	ANN (Model 1)		LSTM 1 layer (Model 2)		CNN 2 layer with LSTM 1 layer (Model 3)		Optimized CNN 2 layer with LSTM 1 layer (Model 4)		CNN 4 layer with LSTM 2 layer (Model 5)	
		R	W	R	W	R	W	R	W	R	W
1	WALKING	447	49	455	41	476	20	471	25	471	25
2	WALKING_UPSTAIRS	396	75	468	3	440	31	455	16	462	9
3	WALKING_DOWNSTAIRS	381	39	420	0	409	11	415	5	416	4
4	SITTING	333	158	375	116	377	114	420	71	405	86
5	STANDING	453	79	428	104	486	46	451	81	453	79
6	LAYING	537	0	534	3	510	27	537	0	522	15
	Overall R/W	2547	400	2680	267	2698	249	2749	198	2729	218



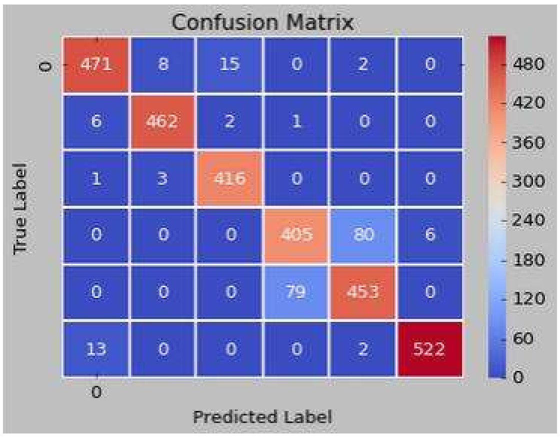


(a) ANN (Model 1) (b) LSTM 1 layer (Model 2)





(c) CNN 2 layer with LSTM 1 layer (Model 3) (d) Optimized CNN 2 layers- LSTM 1 (Model 4)



(e) CNN 4 layers-LSTM 2 layers.

Figure 13: Confusion Matrix for all five models.

From the above results, we determine train accuracy and validation accuracy conducted in different methodologies and observe their variations. Here we observe that different methodology provides different test results depending on their neural network. Figure 13 and Table 4 show class-wise right and wrong predictions over four different models. From Table 2, we see that 2749 samples are correctly classified and only 198 samples are misclassified. From the dataset we used to run this procedure we find that the Optimized CNN 2 layer with LSTM 1 layer (Model 4) provides better output than the other models.

4. Conclusion

In this survey, we carried out a comprehensive study of state-of-the-art methods of human activity recognition. In this paper, we have successfully implemented five ANN, convolutional neural networks, and deep learning models and analyzed the performance using the UCI HAR smartphone dataset. We represented different methods for HAR and computed their accuracy towards achieving the best output and compared their accuracy to determine relatively the methods for human activity recognition. The obtained experimental result reveals that the Optimized CNN 2 layer with LSTM 1 layer (Model 4) provides better output than the other models. Our future work aims at operating with various datasets and more reliable methods and determining their performance variation with the dataset.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

References

[1] Ronao C.A. & Cho SB. (2015). Deep Convolutional Neural Networks for Human Activity Recognition with Smartphone Sensors. In: Arik S., Huang T., Lai W., Liu Q. (eds) Neural Information Processing. ICONIP 2015. Lecture Notes in Computer Science, vol 9492. Springer, Cham. https://doi.org/10.1007/978-3-319-26561-2_6

[2] Moya Reuda, F., Grzezick, R., A. Fink, G., Feldhorst, S., & Ten Hompel, M. (2018). Convolutional Neural Networks for Human Activity Recognition Using Body-Worn Sensors, Journal of Informatics (5020026). https://doi.org/10.3390/informatics5020026

[3] Jobanputra, C., Bavishi, J., & Doshi, N. (2019). Human activity recognition: A survey. *Procedia Computer Science*, *155*, 698-703. https://doi.org/10.1016/j.procs.2019.08.100

[4] Xu, W., Pang, Y., Yang, Y., & Lui, Y. (2018). Human Activity Recognition Based on Convolutional Neural Network, (2018), 24th International Conference on Pattern Recognition (ICPR) Beijing, China, August 20-24. https://doi.org/10.1109/ICPR.2018.8545435

[5] A. Jain and V. Kanhangad, (2018). Human Activity Classification in Smartphones Using Accelerometer and Gyroscope Sensors, IEEE Sensors Journal, 18(3), 1169-1177. https://doi.org/10.1109/JSEN.2017.2782492

[6] Yang, J., Nguyen, M. N., San, P. P., Li, X. L., & Krishnaswamy, S. (2015). Deep convolutional neural networks on multichannel time series for human activity recognition. In *Twenty-fourth international joint conference on artificial intelligence (IJCAI), 3995-4001*.

[7] Wang, H., Zhao, J., Li, J., Tian, L., Tu, P., Cao, T., & Li, S. (2020). Wearable sensor-based human activity recognition using hybrid deep learning techniques. *Security and communication Networks*, *2020*. https://doi.org/10.1155/2020/2132138

[8] Xi, R., Hou, M., Fu, M., Qu, H., & Liu, D. (2018). Deep dilated convolution on multimodality time series for human activity recognition. In *2018 international joint conference on neural networks (IJCNN)*, 1-8. https://doi.org/10.1109/IJCNN.2018.8489540

[9] Ignatov Andrey (2018). Real-time human activity recognition from accelerometer data using Convolutional Neural Networks, Applied Soft Computing, Volume 62, 2018, Pages 915-922, ISSN 1568-4946, https://doi.org/10.1016/j.asoc.2017.09.027

[10] Alake, R (2020). Batch Normalization in Neural Networks Explained (Algorithm Breakdown) https://towardsdatascience.com/batch-normalization-in-neural-networks-1ac91516821

[11] Dua, D. & Graff, C. (2019). UCI Machine Learning Repository. Irvine, CA: University of California, School of Information and Computer Science. http://archive.ics.uci.edu/ml

[12] Hammerla, N. Y., Halloran, S., & Plötz, T. (2016). Deep, convolutional, and recurrent models for human activity recognition using wearables. *arXiv preprint arXiv:1604.08880*. https://doi.org/10.48550/arXiv.1604.08880

[13] Song-Mi Lee, Sang Min Yoon, and Heeryon Cho. (2017). Human activity recognition from accelerometer data using Convolutional Neural Network. 2017 IEEE International Conference on Big Data and Smart Computing (BigComp), 2017, 131-134. https://doi.org/10.1109/BIGCOMP.2017.7881728

[14] Jayabalan, A., Karunakaran, H., Murlidharan, S., & Shizume, T. (2016). Dynamic Action Recognition: A convolutional neural network model for temporally organized joint location data. https://doi.org/10.48550/arXiv.1612.06703.